We claim:

1. A method for generating a visitor segmentation for a visitor population having a plurality of visitors in response to the performance of different messages for different visitors; said method including:

reading message performance results representing message trials and message successes for a message from at least one previous stage in a multi-stage messaging campaign;

computing a current message state on the basis of said prior stage message state and said message performance results; and

generating a visitor segmentation based on said computed current message state and using (i) visitor profile data, (ii) user-defined campaign constraints, and (iii) user-defined campaign objectives.

2. The method in claim 1, wherein said current message state is computed for a cell, and said cell corresponds to a value of a message attribute.

3. The method in claim 2, wherein said cell comprises a single attribute cell or a multi-attribute cell.

4. The method in claim 2, further comprising deriving a message allocation for each cell based on said segmentation.

5. The method in claim 2, further comprising reading prior stage message state for a cell pertaining to a prior stage in a message campaign.

6. The method in claim 2, further comprising:

reading prior stage message state for a cell pertaining to a prior stage in a message campaign prior to said reading performance data; and

deriving a message allocation for each cell based on said segmentation after generating said segmentation.

7. The method in claim 1, further comprising: storing the results of the derived message allocation as a starting point for a next iteration of said method.

8. The method in claim 2, further comprising: storing the results of the derived message allocation as a starting point for a next iteration of said method.

9. The method in claim 1, further comprising repeating said reading, computing, and generating using said derived message allocation for a next iteration of said method.

10. The method in claim 8, where said message is selected from the group consisting of: a banner ad, an email, a pop-up window, a web page, a web layout, a survey, a sales promotion, a political campaign, a public service announcement, a splash page, and combinations thereof.

11. The method in claim 1, where said message is selected from the group consisting of: a banner ad, an email, a pop-up window, a web page, a web layout, a survey, a sales promotion, a political campaign, a public service announcement, a splash page, and combinations thereof.

12. The method in claim 3, where said message is selected from the group consisting of: a banner ad, an email, a pop-up window, a web page, a web layout, a survey, a sales promotion, a political campaign, a public service announcement, a splash page, and combinations thereof.

13. The method in claim 12, wherein said data on the performance of the messages for each cell is selected from the set of data consisting of: selected earlier performance data; all earlier performance data; only the earlier performance data collected after a fixed date; all earlier performance data discounted by a factor; selected earlier performance data discounted by a factor; all earlier performance data discounted by a discounting function; selected earlier performance data discounted

by a discounting function; all earlier performance data, which data may include trials and successes, and which data is discounted by a general weight function.

14. The method in claim 13, wherein said data on the performance of the messages includes prior judgment or information.

15. The method in claim 14, wherein each cell is represented by at least one cell attribute.

16. The method in claim 14, wherein each cell is represented by a plurality of cell attributes.

17. The method in claim 16, wherein each attribute corresponds to one or more subspaces of a space consisting of a defined set of possible visitors.

18. The method in claim 17, wherein said defined set of possible visitors comprises all actual visitors.

19. The method in claim 17, wherein said defined set of possible visitors comprises all possible visitors.

20. The method in claim 17, wherein each cell corresponds to an attribute value along one of the visitor attributes or a combination of attribute values along a subset of the visitor profile information.

21. The method in claim 20, wherein said visitor information attribute is selected from the group consisting of: a number of times that any visitor has been shown each message in the past; visitor profiling data collected during the current visit; visitor profiling data collected during a prior visit; visitor profiling data collected from an external source; and combinations thereof.

22. The method in claim 20, wherein said visitor information attribute includes visitor profiling data collected during the current visit including at least one of the following: time of day, IP address, geographical location, Internet service provider, browser type, computer system type, and the web site upon which the message is viewed.

23. The method in claim 20, wherein said visitor information attribute includes the visitor profiling data collected during a prior visit including at least one of the following: number of previous responses to each message alternative, identification of web sites upon which the visitor has viewed messages in the past, and a quality or extent of visitor purchase or participation as a result of previous interactions with each message alternative.

24. The method in claim 20, wherein said visitor information attribute includes the visitor profiling data collected during a prior visit and said visitor profiling data collected during the prior visit corresponds to demographic or psychographic information about the visitor.

25. The method in claim 24, wherein said demographic or psychographic information about the visitor is selected from the set consisting of: age, gender, income, geographic location, consumer preferences, purchasing habits, interests, hobbies, and combinations thereof.

26. The method in claim 24, wherein said stage has a stage duration and said stage duration is a period of time.

27. The method in claim 24, wherein said stage has a stage duration, and said stage duration is a period of time selected from the set of period of times consisting of: less than 1 second, a 1 second to 60 second long period, a 1 minute to 1 hour long period, a 1 hour to 3 hour long period, a 3 hour to 9 hour period, a 9 hour to 15 hour period, a 15 hour to 24 hour long period, a 1 hour to 24 hour time period, a 2 hour to

72 hour period, a one week period, a period of time beginning with a first defined event and terminating with a second defined event, and combinations thereof.

28. The method in claim 2, wherein said campaign constraints include at least one of the following: (i) a message may not be shown at some cells or zones, (ii) a message may not be shown during some stages, (iii) a message is deleted from a zone at some stage in the campaign, (iv) a message is added to a zone at some stage in the campaign, (v) a zone is deleted at some stage of the campaign, (vi) a zone is added at some stage of the campaign, (vii) a limitation as to when or where a message may be shown, (viii) a requirement as to when or where a message must be shown, and combinations thereof.

29. The method in claim 4, wherein said message allocation is computed using one or more of the following procedures (i-iv), where when more than one of the procedures are used, the procedures may be applied in any order:

(i) a segmentation procedure to determine multiple segments for each of one or more templates;

(ii) a template-choosing procedure for determining the best template from a set of possible templates for segmentation;

(iii) a procedure pooling performance information for a cell in the best template from all the cells belonging to the same segment in the best template; and

(iv) an allocation procedure to determine an allocation vector for visitors corresponding to each cell of the best template, said allocation vector specifying a fraction of visitors that will be shown each of the message alternatives for the future single stage or for future multiple stages; and

wherein said templates each comprise a collection grid of one or more cells, and the cells included within different templates may be the same or different from the cells included in another template.

30. The method in claim 29, wherein when the allocation procedure to determine an allocation vector for visitors corresponding to each cell of the best template is used, the allocation procedure uses pooled performance information.

31. The method in claim 29, wherein when the allocation procedure to determine an allocation vector for visitors corresponding to each cell of the best template is used, the allocation procedure does not use pooled performance information.

32. The method in claim 29, wherein said pooling performance information includes information summing the total number of discounted impressions and clicks for all the cells belonging to a segment to determine the pooled impressions and clicks respectively for a cell.

33. The method in claim 29, wherein the segmentation procedure comprises a segmentation procedure comprising: an iterative segmentation procedure comprising at least one iteration of (i) a procedure for choosing a subset of segmentations, where the subset changes from one iteration to another iteration when multiple iterations are executed, and (ii) a procedure for comparing these segmentations to find the best current segmentation.

34. The method in claim 29, wherein the segmentation procedure comprises a segmentation procedure comprising: a recursive segmentation procedure where at every step of the recursion, the segmentation procedure decides whether to split a segment into two or more disjoint smaller segments or not to split a segment, and wherein splitting is deemed better if the weighted performance of all the disjoint smaller segments is better than the entire unsplit segment.

35. The method in claim 33, wherein the best current segmentation is determined from the set of determinations consisting of: (a) determining the best segmentation based on estimating the performance of each segmentation; (b) determining the segmentation which best models the underlying performance; (c) determining the segmentation which best models the underlying performance and uses the Akaike Information Criterion for modeling the relative performance; (d) determining the segmentation which best models the underlying performance and uses the

Likelihood Ratio Test for modeling the relative performance; (e) determining the segmentation which best models the underlying performance and uses the Likelihood Ratio Test using Wilkes Hypothesis testing for modeling the relative performance; and (f) determining the segmentation which best models the underlying performance but does not use the Akaike Information Criterion or the Likelihood Ratio Test for modeling the relative performance.

36. The method in claim 35, wherein said performance is measured using a performance measure determined by a performance measurement procedure comprising:

for each segment, determining θ, where θ is the largest value for which P(click-through rate for at least one banner > θ) = 0.95;

estimating the performance of a segmentation as the average, weighted by the number of impressions, of the performance of each segment within the segmentation; and

identifying the segmentation with the highest estimated performance as the best segmentation.

37. The method in claim 30, wherein said determining the segmentation which best models the underlying performance uses the Akaike Information Criterion, and said Akaike Information Criterion comprises:

computing Akaike's Information Criterion (AIC) for each segmentation g, using the relationship:

$$AIC(g) = -2 \log(L(\theta | x)) + 2K$$

where θ is the maximum likelihood estimator of the parameter vector of the segmentation g containing the performance estimate for each segment, x is an observation of the underlying random variable, L is the likelihood function of θ given the data, and K is the number of parameters in the segmentation g or the size of the vector θ; and

identifying the segmentation with the smallest value of AIC as being the best segmentation.

38. The method in claim 35, wherein said Likelihood Ratio Test for modeling performance comprises:

defining Λ as -2 log (L_0 / L_1), where L_0 is the likelihood of a null model with n parameters, evaluated with the maximum likelihood parameter values, and L_1 is the likelihood of an alternative model with $m > n$ parameters, evaluated at its maximum likelihood parameter values, such that Λ is asymptotically distributed as χ^2_{m-n} by Wilk's theorem; and

identifying a first segmentation associated with L_1 when Λ exceeds a value corresponding to a particular p-value as being a better segmentation than a second segmentation associated with L_0 .

39. The method in claim 38, wherein said particular p-value is chosen to be in the range between substantially 0.5 and 0.7.

40. The method in claim 38, wherein said particular p-value is chosen to be in the range between substantially 0.7 and 0.9.

41. The method in claim 38, wherein said particular p-value is chosen to be in the range between substantially 0.1 and 1.0.

42. The method in claim 38, wherein said particular p-value is chosen to be in the range between substantially 0.9 and 1.0.

43. The method in claim 38, wherein said particular p-value is chosen to be in the range between substantially 0.9 and 0.99.

44. The method in claim 38, wherein said particular p-value is chosen to be 0.9, 0.95, 0.99, and any intermediate value there between.

45. The method in claim 38, wherein said subset of segmentations chosen depends on or is biased by one or more of the following: (a) a prior segmentation provided by

the user; (b) a skeleton segmentation constraint provided by the user; and (c) a segmentation determined at the end of previous stage.

46. The method of claim 45, wherein when said subset of segmentations chosen depends on or is biased by said prior segmentation provided by the user, said prior segmentation may optionally be used to give preferential treatment to a subset of all possible segmentations.

47. The method of claim 45, wherein when said subset of segmentations chosen depends on or is biased by said skeleton segmentation constraint provided by the user, then the segmentation chosen is such that each segment of the segmentation is a proper subset of one of the segments of the skeleton segmentation provided.

48. The method of claim 45, wherein when said subset of segmentations chosen depends on or is biased by said segmentation determined at the end of previous stage, then the segmentation chosen is such that each segment of the segmentation is a proper subset of one of the segments of the segmentation used in the previous stage.

49. The method in claim 45, wherein the best template is chosen based on estimating the performance of the best segmentation for each template.

50. The method in claim 49, wherein the allocation vector is derived using the Pairwise Algorithm.

51. The method in claim 50, wherein a recursive greedy procedure is used to determine the segmentation for a given template; said procedure including a decision performed at each repetition of the recursive procedure as to whether to split a parent set into two children along one attribute of a given template.

52. The method in claim 51, wherein said decision performed at each repetition of the recursive procedure as to whether to split a parent set into two children along one

attribute of a given template includes performing the following sub-procedure comprising:

deriving performance information for a supercell by pooling all of the cells corresponding to said supercell, the supercells along a given attribute consisting of all of the cells with the same value of the attribute;

finding the attribute and split that results in the best segmentation; and

if a split leads to a favorable segmentation as compared to not segmentation, the split of the parent into two children is made.

53. The method in claim 52, further comprising recursively applying said procedure to each of the two children.

54. The method in claim 52, wherein said finding the attribute and split that results in the best segmentation, comprises a procedure that is applied along each attribute to find the attribute and split which results in the best segmentation.

55. The method in claim 54, wherein said procedure finding the attribute and split that results in the best segmentation includes: (i) defining at least one supercell compring a plurality of cells; (ii) beginning with all supercells grouped together as a first set (Set A); (iii) finding the one best supercell that, when split from the rest, leads to the best overall segmentation according to some criteria, and defining a second set (Set B) to contain this one supercell; (iv) finding the next best supercell to remove from said first set (Set A) and adding said removed next best supercell to said second set (Set B); (v) continue, finding said best supercell and said next best supercell until all there are only two supercells left in said first set (Set A); and (vi) choosing the split that gives the highest split value of all combinations tried above.

56. The method in claim 55, wherein said at least one supercell comprises a supercell along a given attribute having a plurality of cells with the same value of the attribute.

57. The method in claim 56, wherein said the performance information of a supercell is derived by pooling performance information for all the cells corresponding to said supercell.

58. The method in claim 53, wherein said next best supercell is identified in terms of best split value amongst the choices.

59. The method in claim 1, wherein said method for segmenting visitors from a population based on the performance of different messages against different visitors is performed with the goal of improving the stage-to-stage performance a multi-stage message campaign.

60. The method in claim 3, wherein said method for segmenting visitors from a population based on the performance of different messages against different visitors is performed with the goal of improving the stage-to-stage performance a multi-stage message campaign.

61. The method in claim 1, wherein said method for segmenting visitors from a population based on the performance of different messages against different visitors is performed without the goal of improving the stage-to-stage performance a multi-stage message campaign.

62. The method in claim 3, wherein said method for segmenting visitors from a population based on the performance of different messages against different visitors is performed without the goal of improving the stage-to-stage performance a multi-stage message campaign.

63. A method for generating a visitor segmentation for a visitor population having a plurality of visitors in response to the performance of different messages for different visitors; said method including:

reading prior stage message state for a cell pertaining to a prior stage in a message campaign

reading message performance results representing message trials and message successes for a message from at least one previous stage in a multi-stage messaging campaign;

computing a current message state for a cell comprising one or a combination of two or more message attributes on the basis of said prior stage message state and said message performance results;

generating a visitor segmentation based on said computed current message state and using (i) visitor profile data, (ii) user-defined campaign constraints, and (iii) user-defined campaign objectives;

deriving a message allocation for each cell based on said segmentation;

storing the results of the derived message allocation as a starting point for a next iteration of the said method;

repeating said reading, computing, and generating using said derived message allocation for a next iteration of said method;

said message is selected from the group consisting of: a banner ad, an email, a pop-up window, a web page, a web layout, a survey, a sales promotion, a political campaign, a public service announcement, a splash page, and combinations thereof.

said data on the performance of the messages for each cell is selected from the set of data consisting of: selected earlier performance data; all earlier performance data; only the earlier performance data collected after a fixed date; all earlier performance data discounted by a factor; selected earlier performance data discounted by a factor; all earlier performance data discounted by a discounting function; selected earlier performance data discounted by a discounting function; all earlier performance data, which data may include trials and successes, and which data is discounted by a general weight function;

said data on the performance of the messages includes prior judgment or information;

each cell is represented by at least one cell attribute;

each attribute corresponds to one or more subspaces of a space consisting of a defined set of possible visitors;

said defined set of possible visitors comprises all possible visitors;

each cell corresponds to an attribute value along one of the visitor attributes or a combination of attribute values along a subset of the visitor profile information;

said visitor information attribute is selected from the group consisting of: a number of times that any visitor corresponding to said cell has been shown each message in the past; visitor profiling data collected during the current visit; visitor profiling data collected during a prior visit; visitor profiling data collected from an external source; and combinations thereof;

said visitor information attribute includes visitor profiling data collected during the current visit including at least one of the following: time of day, IP address, geographical location, Internet service provider, browser type, computer system type, and the web site upon which the message is viewed;

said visitor information attribute includes the visitor profiling data collected during a prior visit including at least one of the following: number of previous responses to each message alternative, identification of web sites upon which the visitor has viewed messages in the past, and a quality or extent of visitor purchase or participation as a result of previous interactions with each message alternative;

said visitor information attribute includes the visitor profiling data collected during a prior visit and said visitor profiling data collected during the prior visit corresponds to demographic or psychographic information about the visitor, said demographic or psychographic information about the visitor is selected from the set consisting of: age, gender, income, geographic location, consumer preferences, purchasing habits, interests, hobbies, and combinations thereof;

said campaign constraints include at least one of the following: (i) a message may not be shown at some cells or zones, (ii) a message may not be shown during some stages, (iii) a message is deleted from a zone at some stage in the campaign, (iv) a message is added to a zone at some stage in the campaign, (v) a zone is deleted at some stage of the campaign, (vi) a zone is added at some stage of the campaign, (vii) a limitation as to when or where a message may be shown, (viii) a requirement as to when or where a message must be shown, and combinations thereof;

said message allocation is computed using one or more of the following procedures, where when more than one procedure is used, the procedures may be applied in any order: (i) a segmentation procedure to determine multiple segments

for each of the templates; (ii) a template-choosing procedure for determining the best template for segmentation; (iii) a procedure pooling performance information for a cell in the best template from all the cells belonging to the same segment in the best template; (iv) an allocation procedure to determine an allocation vector for visitors corresponding to each cell of the best template, said allocation vector specifying a fraction of visitors that will be shown each of the message alternatives for the future single stage or for future multiple stages;

when the allocation procedure to determine an allocation vector for visitors corresponding to each cell of the best template is used, and the allocation procedure optionally uses pooled performance information, said pooling performance information includes information summing the total number of discounted impressions and clicks for all the cells belonging to a segment to determine the pooled impressions and clicks respectively for a cell; and

the segmentation procedure comprises at least one of an iterative segmentation procedure and a recursive segmentation procedure; said iterative segmentation procedure comprising at least one iteration of (i) a procedure for choosing a subset of segmentations, where the subset changes from one iteration to another iteration when multiple iterations are executed, and a procedure for comparing these segmentations to find the best current segmentation; and said recursive segmentation procedure where at every step of the recursion, the segmentation procedure decides whether to split a segment into two or more disjoint smaller segments or not to split a segment, and wherein splitting is deemed better if the weighted performance of all the disjoint smaller segments is better than the entire unsplit segment.

64. The method in claim 63, wherein the best current segmentation is determined from the set of determinations consisting of: (a) determining the best segmentation based on estimating the performance of each segmentation; (b) determining the segmentation which best models the underlying performance; (c) determining the segmentation which best models the underlying performance and uses the Akaike Information Criterion for modeling the relative performance; (d) determining the

segmentation which best models the underlying performance and uses the Likelihood Ratio Test for modeling the relative performance; (e) determining the segmentation which best models the underlying performance and uses the Likelihood Ratio Test using Wilkes Hypothesis testing for modeling the relative performance; and (f) determining the segmentation which best models the underlying performance but does not use the Akaike Information Criterion or the Likelihood Ratio Test for modeling the relative performance.

65. The method in claim 64, wherein said performance is measured using a performance measure determined by a performance measurement procedure comprising:

for each segment, determining θ, where θ is the largest value for which P(click-through rate for at least one banner $> \theta$) = 0.95;

estimating the performance of a segmentation as the average, weighted by the number of impressions, of the performance of each segment within the segmentation; and

identifying the segmentation with the highest estimated performance as the best segmentation.

66. The method in claim 65, wherein said determining the segmentation which best models the underlying performance uses the Akaike Information Criterion, and said Akaike Information Criterion comprises:

computing Akaike's Information Criterion (AIC) for each segmentation g, using the relationship:

$$AIC(g) = -2 \log(L(\theta | x)) + 2K$$

where θ is the maximum likelihood estimator of the parameter vector of the segmentation g containing the performance estimate for each segment, x is an observation of the underlying random variable, L is the likelihood function of θ given the data, and K is the number of parameters in the segmentation g or the size of the vector θ; and

identifying the segmentation with the smallest value of AIC as being the best segmentation; and

said Likelihood Ratio Test for modeling performance comprises:

defining Λ as -2 log (L_0 / L_1), where L_0 is the likelihood of a null model with *n* parameters, evaluated with the maximum likelihood parameter values, and L_1 is the likelihood of an alternative model with $m > n$ parameters, evaluated at its maximum likelihood parameter values, such that Λ is asymptotically distributed as χ^2_{m-n} by Wilk's theorem; and

identifying a first segmentation associated with L_1 when Λ exceeds a value corresponding to a particular p-value as being a better segmentation than a second segmentation associated with L_0 ; and

said subset of segmentations chosen depends on or is biased by one or more of the following: (a) a prior segmentation provided by the user; (b) a skeleton segmentation constraint provided by the user; and (c) a segmentation determined at the end of previous stage.

67. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism, comprising: a program module that directs the computer system and/or components thereof, to function in a specified manner to generate a visitor segmentation for a visitor population having a plurality of visitors in response to the performance of different messages for different visitors, the program module including instructions for:

reading message performance results representing message trials and message successes for a message from at least one previous stage in a multi-stage messaging campaign;

computing a current message state on the basis of said prior stage message state and said message performance results; and

generating a visitor segmentation based on said computed current message state and using (i) visitor profile data, (ii) user-defined campaign constraints, and (iii) user-defined campaign objectives.

68. The computer program product in claim 67, wherein said current message state is computed for a cell, and said cell comprises one or more message attributes.

69. A method of visualizing the performance of a set of messages, said method characterized in that said method includes virtual or actual construction of a two-dimensional graphical plot, and further characterized in that the x-axis of the graphical plot runs over the different messages, the y-axis of the graphical plot runs over different levels or values of a message attribute, and the color or shade of gray represents the performance of that message for that level or value of the targeting attribute.

70. A method of visualizing the performance of a set of messages, said method characterized in that said method includes virtual or actual construction of a two-dimensional graphical plot, and further characterized in that the x-axis of the graphical plot runs over one targeting attribute, the y-axis of the graphical plot runs over another targeting attribute, and the color or shade of gray represents the performance of the best message for that attribute combination.

71. A method of visualizing the performance of a set of messages, said method characterized in that said method includes virtual or actual construction of a two-dimensional graphical plot, and further characterized in that the x-axis of the graphical plot runs over one targeting attribute, the y-axis of the graphical plot runs over another targeting attribute, and the color or shade of gray identifies the message that performs best for that attribute combination.

72. A method for improving the performance of advertising messages in an interactive measurable medium, said method comprising:

dividing visitors into segments based on the performance of different messages for different visitors;

constructing a segmentation, based on the performance of message alternatives for different visitor types, where visitors are classified by demographic or psychographic information along a number of attributes; and

comparing different segmentations to determine which segmentation will lead to the maximum visitor response.

73. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism, comprising: a program module that directs the computer system and/or components thereof, to function in a specified manner to improve the performance of advertising messages in an interactive measurable medium, the program module including instructions for:

dividing visitors into segments based on the performance of different messages for different visitors;

constructing a segmentation, based on the performance of message alternatives for different visitor types, where visitors are classified by demographic or psychographic information along a number of attributes; and

comparing different segmentations to determine which segmentation will lead to the maximum visitor response.